新冠病毒独立自然连续传播系统的建立及简单模型验证

王基庆* 李志国
（华东师范大学物理与电子科学学院 上海 200241）

摘要：
[目的] 验证新冠病毒输入I代到感染II代之间独立自然连续传播理想体系；分析非湖北地区疫情特点和走势。
[方法] 使用正态分布的线性叠加拟合全国除湖北外每日新增的数据。
[结果] 拟合计算得到的新冠病毒可再生系数$R_0 = 2.2$，一代到二代的峰值传播时间是7天。
[局限] 病毒携带者流入和管控时间差要在传播峰值内。
[结论] 湖北外新冠病毒发展趋势近似满足独立自然连续传播理想系统。

关键词：新冠病毒；可再生系数；理想体系

1. 引言和模型原理

2019 年 12 月份以来的新冠病毒肆虐引起了世界各国政府和科学家的关注[1-3]。为了控制病毒的进一步蔓延，武汉市政府从 1 月 23 日开始采取严格的管控措施，管控的目的，从物理空间上讲就是：让湖北成为孤立系统，外省不和湖北有交换。湖北外也做到一级响应，有症状或疑似全部隔离，这样保证了湖北外主要是在输入 I 代和感染 II 代之间有传播，II 代以上可能性较低。有了这两个条件，就保证了湖北外地区是个近似的只有 I、II 代感染的理想系统。这样求出的病毒基本可再生系数 R_0 才有意义，没有隔代感染混淆。

另外从时间上看，由于春运，湖北出省高峰大约在 1 月 17 日(小年夜)附近，这些携带者到了外地变成了一代传播者，各地一级响应在 23 日以后，也就是说大概七天后，外省比较明显地采取了干预措施，在这之间散发到外省的病毒携带者基本上是自然状态(无干预)从输入 I 代到感染 II 代，而这七天正好是一带症状显著以及带间的自然传染一个时间峰值窗口，再往后的话，由于各地干预，各种发热或者疑似以及密切接触者多被很好地被收治或隔离，II 代到 III 代的感染可能性比较低。

上述在空间和时间上对体系的分割就保证了目前我们所考虑的非湖北地域基本满足和湖北无交换、无隔代感染、一个周期内无干预条件下连续自然传播理想系统。在这种情况下，除湖北外全国每日新增冠状病毒人目基本上输入 I 代和感染 II 代新增的总和。这样求得的 I 代传给 II 代的自然状态下感染率应符合实际情况。基于 SEIR 传播动力学原理中感染者的正态分布特征[4,5]，我们应用简单线性叠加模型拟合除湖北外全国每日新增冠状病毒疫情图(数据截止到 2.19 日，

数据来源于凤凰网和各地卫生和健康委员会网站，后续要统一到世界卫生组织和各地卫生和健康网站上的权威发布上。由于不同来源的数据差别不大，不影响对结果的分析和判断上)，见下图 1：

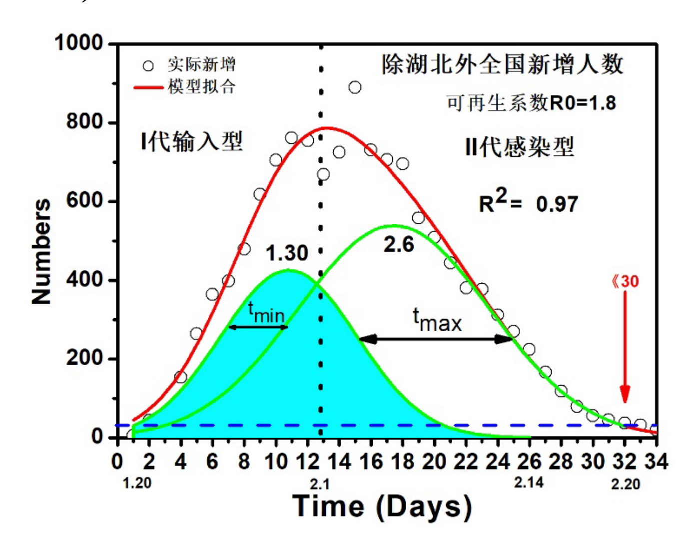


图 1 除湖北外全国实际每日新增数据和拟合结果

图中空心圆点表示每日实际新增，红色曲线是模型拟合结果，拟合准确度达到了 97%。从图中可以看出，模型很好拟合了实际新增的发展趋势。我们模型建在 2 月 13 日，之后七日我们预测和当日实际新增见以下表格：

日期	实际新增	预测	误差(%)
13	270	249	7.8
14	224	200	11.2
15	166	156	6.0
16	118	118	0
17	79	87	-10.1
18	56	63	-12.5
19	45	44	2

从表格中看，误差基本在 10%附近，在统计误差范围之内，说明现在每日新增的对象(非湖北地区)可以近似看成一个独立的只有 I-II 代感染的非干预的系统。因此这种模型方法测得的 R_0 值应能很接近自然状态下的真实值。

2. 结果和讨论

2.1 结果分析

基于图 1，我们得出如下结论和预测：

(1) 2 月 1 号前外省以输入型新增为主，之后以 II 代感染型为主；

(2) 湖北外输入 I 代传给感染 II 代的平均感染率(可再生系数)R_0 = 1.8，如去掉干扰因素 R_0 = 2.2 (见下面讨论和图 3)。和很多文献报道一致[6-8]；

(3) 平均最短发病窗口期 t_{min} = 4 天，最长窗口期 t_{max} = 10 天，高峰窗口期 7 天；

(4) 外推湖北春运出省高峰在小年夜前后，除湖北外全国疫情在 20 号左右降低到 30 附近(除湖北外平均每省一个)；

(5) 春运返程对疫情影响有限，因为各地严格控制了 I 代、II 代、疑似及密切接

触者的流动。

2.2 进一步深入分析：

(1)从图 2 我们可以看出有两个实际新增数据点偏离理论预测值较大，其绝对和相对误差分别高达 A：118(-17.5%)，B：128(14.5%),相对误差明显高于系统平均值(10%附近)。如果不加分析认为是系统误差偏离较大或者简单去掉这两点的话，可能会疏忽掉重要信息。

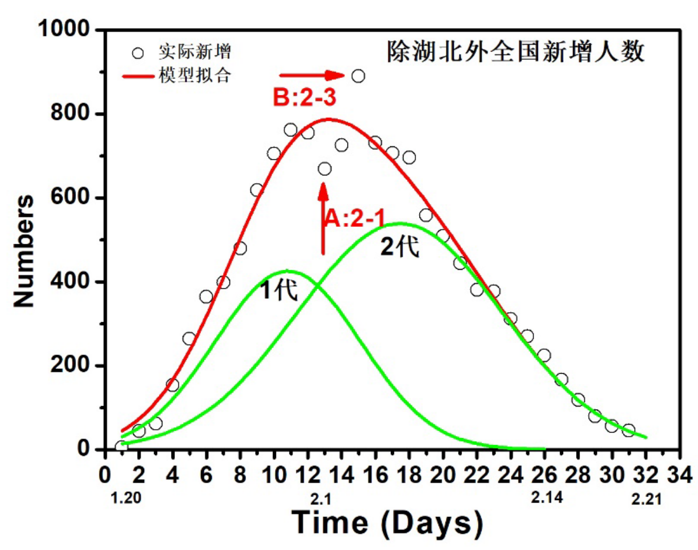


图 2 除湖北外全国实际每日新增数据和拟合结果

①仔细分析 A 点时间我们可以发现：A 点的 2 月 1 日往后追溯七天正好是各地管控开始时间(23 日左右)，可能也由于第二天是农历除夕，传统的团圆习俗让 I 代和 II 代间的感染在这个时间点有明显的减少，这个交叉传染的减少并不影响一代的确诊数，但会导致 I-II 代间感染机会减少，后果是七天后(一个峰值周期)，实际新增值明显地减少。理论上如果没有这个 23 日干预点存在的话，2 月 1 日正好是一个一代到二代新增属性改变的平台期(见红色曲线和 I、II 代绿色曲线的交叉点)。而之前干预时间节点正好降低了平台处（2 月 1 日）的新增值，这点数值上的变化反映了各级政府部门干预的效果。②仔细分析 B 点我们可以发现：B 点在 2 月 3 号，这点新增值偏离较大可能来源于 1 月 10 日春运刚开始出省的一部分小高峰，这部分携带者在湖北内的传染代数应低于 17 号的出省携带者(父代)，其到了外省的传染率$R' < R_0$，经过 10 天的增长期(根据 17 号后 I 代的绿色曲线推算)，其输入 I 代的确诊高峰期可能在 19 号左右。经过一个峰值周期其 II 代在 26 号左右，此时其 I-II 代间传染率R'_{12}应大于原始的 R_0 ($R'_{12} \cong R_0 > R'$)。但是由于其感染总量远小于 17 号出省的这部分值且传播力相当，所以在曲线上并没有明显反映出来。但是当再过一个峰值期(2 月 3 号附近)，可能存在 II-III 代间传染特性，其传播率大幅增长$R'_{23} > R'_{12}$)，这部分倍增系数的提高会导致在这个时间点实际新增值明显大于理论曲线。由于其总量仍然不高，其对曲线的影响只集中在一个时间点上，系统整体还是以 17 号后出省病毒携带者 I-II 代间感染为主。但从上述推断我们估计病毒的传播力可能在增强。

综上对上述两个数据点推断我们认为：这些小的干扰因素并不会影响 17 号后非湖北地区理想的孤立系统状态，在这系统中 I-II 代间相互感染作用占主导地位，因此导致实际新增值和理论曲线高度吻合。但不应忽视这些偏差较大数据点的内在病毒传播动力学机制。

(2)单个省份的分析：从浙江省的每日新增图我们可以看出(图 4,删除部分离散较大数据)，1 月 28-30 日的确诊病例突然增加，由于缺少流行病学调查和当时病例检测数据，尚不清楚其中的原因。考虑到浙江温州在湖北从事商业的人较多，如果数据能反映客观情况的话，可能原因是从湖北返回家乡的人集中在之前某几日。例外从拟合的结果可以发现，有些 I 代输入病例在未确诊之前(潜伏期)就有很强的传染性。这方面需要引起足够的重视。从湖南的数据可以看出(图 5)，确诊比例中存在大约 17%的 III 代感染。由于湖南离湖北较近，人员交往密集导致出湖北的人流跨度时间比较长，引起 III 代感染占有一定的比例。类似的情况也发现在安徽、江西和河南等周边省份。

(3) 2 月 20 日发生在山东和浙江两省监狱区新增病例没有被放入到我们的统计数据中，由于流行病学调查显示这些病例是由 II 代以上的感染引起的，没有放入我们的统计中应是合理。同时也暗示由于存在潜在的变异，新冠病毒 II 代以上的感染率不容忽视，特别是在人员积聚区域。

2.3 和流行病学调查对比：

目前各地关于确诊病例流行病学溯源数据较少，但是上海市卫健会每日通报中包括这方面的调查。截至 2 月 23 日，确诊 335 病例中，144 例有湖北居住或旅行史，33 例有湖北以外地区居住或旅行史,158 例有相关病例接触史，据此我们推算的传播率是 1.3。另外山东在 2 月 9 日通报中山东 379 例确诊病例中 153 例来自湖北，传播率是 1.5.进一步的数据对照我们会随时跟进。

2.4 模型优点：

(1) 依据当前系统进行预测，能做到简单、快速、准确；

(2) 模型简单，但却是可以用来验证系统是否是自然、非干预和连续感染的理想模型，可确保计算可再生系数对象满足理想系统条件；

(3) 可以增加干预度 C 参数来观察干预度后某一时段内的效果；

(4) 三代内，对适当超出 I-II 代间的传播也能做到预测。

2.5 模型适应条件:

II 代内传播为主，独立非干预系统。1 月 20 日后外省散发并且之后确诊和收治及时，保证了模型条件基本符合。

2.6 误差来源：

(1) 在春运出省高峰期前，部分病毒携带者流入到外省，可能会导致少许 II-III 带间感染，见上述深入讨论部分；

(2) 各地干预期和发病或感染窗口期的偏差可能会引入部分非自然因素下的病毒可再生系数。但是春运的巧合(一般小年夜前后出行高峰)以及七天后(一个峰值周期内)的湖北封省及各地严格管控使得现在实际新增数据正好反映了输入 I 代和感染 II 代间自然感染的理想状态。因此即使使用非常简单方法也能得到接近真实的自然传染率数据结果;

(3) 检测压力引起的确诊病例滞后。

另外从全国大数据样本看，上述这三个误差来源由于统计样本多，根据统计学原理，其对真实数据的影响不会很大。

3. 结论与启示

通过上述分析，我们得到如下启示：

(1) 抓住问题主要矛盾(主要相互作用项)，忽略次要的干扰因素(微扰项)，从而使问题简化是物理学处理复杂体系常用方法。越是简单理想系统，所用模型方法也越简单。所谓简单就是美-Simple is Beautiful；

(2) 对于偏离理论值比较大的数据点不要轻易删除或放弃，其中可能存在微扰项的作用，见上述深入讨论部分；

(3) 实际值和预测曲线的高度吻合说明除湖北外各地确诊数据及时、准确上报，以及信息公开透明机制的重要性，而这对以后的公共危机预防、控制和新冠病毒流行病学研究至关重要。

参考文献

[1] Chan J F, Yuan S, Kok K, et al. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster[J]. The Lancet, 2020.

[2] Benvenuto D, Giovannetti M, Ciccozzi A, et al. The 2019-new coronavirus epidemic: evidence for virus evolution [J]. Journal of Medical Virology, 2020.

[3] Zhu N, Zhang D Y, et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. N Engl J Med 2020; 382:727-733

[4] Tianmu Chen, Jia Rui1, et al. A mathematical model for simulating the transmission of Wuhannovel Coronavirus. https://doi.org/10.1101/2020.01.19.911669doi: bioRxiv preprint

[5] Luz P M , Struchiner C J , Galvani A P , et al. Modeling Transmission Dynamics and Control of Vector-Borne Neglected Tropical Diseases[J]. PLoS Neglected Tropical Diseases, 2010, 4(10):e761

[6] Zhao S, Lin Q Y, et al., Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: A data-driven analysis in the early phase of the outbreak https://doi.org/10.1101/2020.01.23.916395 doi: bioRxiv preprint

[7] Riou J, Althaus CL: Pattern of early human-to-human transmission of Wuhan 2019-nCoV. bioRxiv 2020:2020.2001.2023.917351.

[8] Read J M, Bridgen J R, Cummings D A, Ho A, Jewell CP: Novel coronavirus 2019-nCoV: early estimation of epidemiological parameters and epidemic predictions. medRxiv 2020:2020.2001.2023.20018549.

(通讯作者:王基庆 E-mail:jqwang@ee.ecnu.edu.cn)

作者贡献声明：

王基庆：研究方案和模型总体设计；

李志国：数据采集、整理和分析；

附件：

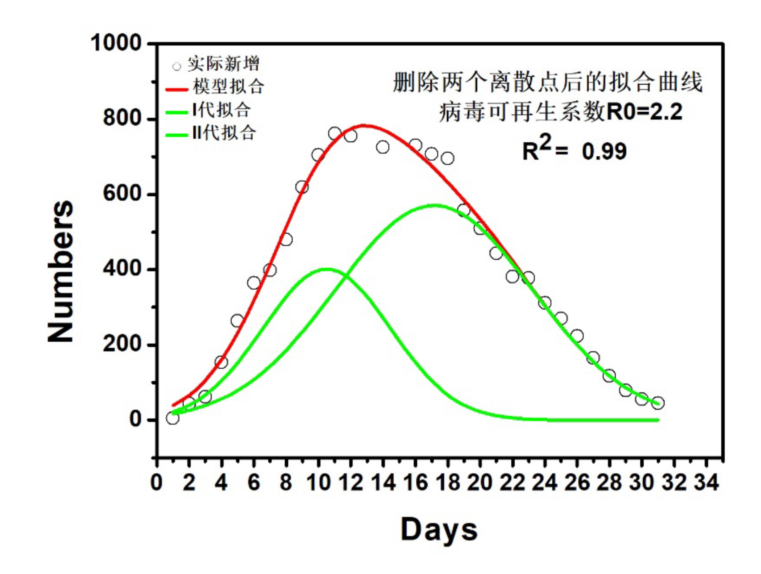


图 3 去掉原图 1 中 A 和 B 两点后的拟合曲线和数据

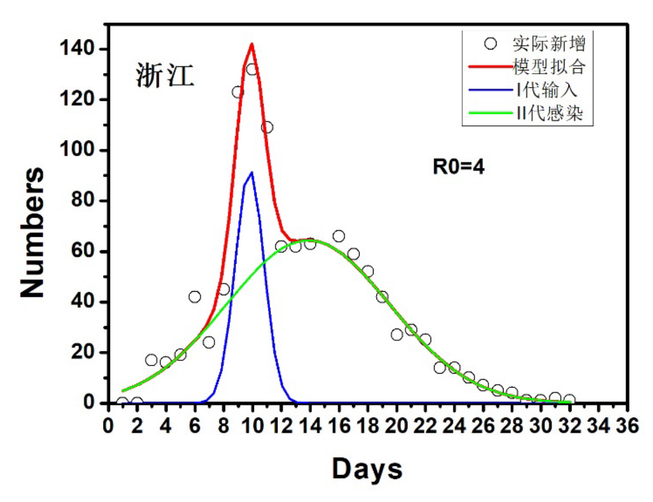


图 4 浙江省实际每日新增数据和拟合结果

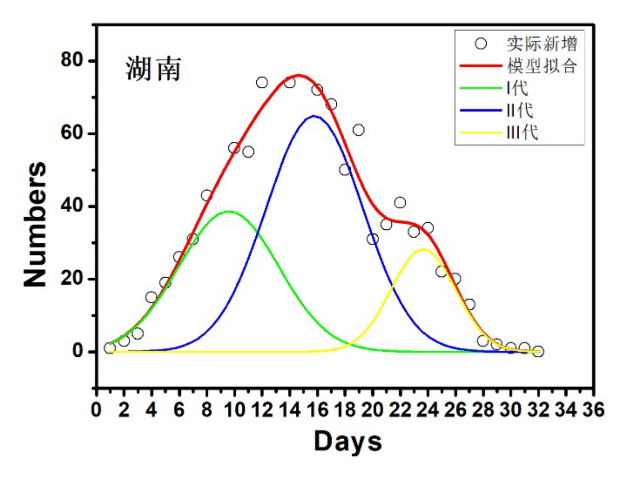


图 5 湖南省实际每日新增数据和拟合结果